Chronological Categorization and Decomposition of Customer Loads

Ghavameddin Nourbakhsh, Gary Eden, Dylan McVeigh, *and* Arindam Ghosh, *Fellow, IEEE*

Abstract— **Majority of distribution utilities do not have accurate information on the constituents of their loads. This information is very useful in managing and planning the network, adequately and economically. Customer loads are normally categorised in three main sectors; residential, industrial and commercial. In this paper, penalized least squares regression and Euclidean distance methods are developed for this application to identify and quantify the makeup of a feeder load with unknown sectors/sub-sectors. This process is done on monthly bases to account for seasonal and other load changes. The error between the actual and estimated load profiles are used a benchmark of accuracy. This approach has shown to be accurate in identifying customer types in unknown load profiles, and is also used in cross-validation of the results and initial assumptions.**

Index Terms— **Load Profiling, Decomposition, Classification, Clustering, K-means**

I. Introduction

As the demand for electricity is increasing, distribution companies who purchase block power for their customers try to efficiently manage their system. For example, diversify their peak power consumption. Reducing peak power has numerous advantages for the distributor, examples of which include saving money in block purchase and reducing the infrastructure costs accounting for possible high loading on transformers.

One measure of reducing peak load is by offering incentives to customer sectors; targeting those which contribute highly to the peak load, in an attempt to change behavioral patterns. This is facilitated by identifying the composition of loads in terms of the amount of power different customer sectors consume.

Smart meters and communication systems can provide detailed information on loads. However, this technology is in its infancy in application to distribution systems. This paper looks into methods where detailed load information can be extracted without the expense of large infrastructure costs and data collection, especially in third world countries where power system funding is limited.

Previous published research works on load profiling incorporate different types of clustering, including K-means [1], fuzzy clustering [2] and two-stage fuzzy clustering [3], all of which are valid methods. Reference [4] also classifies consumers into three main control categories; industrial, commercial and households using both fuzzy and k-means clustering. Studies have also been conducted using distance measures in the clustering algorithm that are a slightly modified form of the Euclidean distance [5]. The distinguishable difference in load profiles of working days and weekend days was analyzed using various clustering techniques in [6]. Reference [7] suggests temperature, and the usage of appliances such as air-conditioners, refrigerators and freezers (particularly industrial type) have a major impact on a load profile, supporting the claim of seasonal variation. The influence of temperature on the load profile has also been observed by means of stochastic load profiling [8].

The research works cited in literature are generally associated with load profile modeling and estimation of different clusters. The lack of quantitative methods for decomposing unknown load profile types in terms of main known customer sectors is the research gap that this paper has identified and has provided methods and evaluations.

By being able to decompose unknown profiles in terms of known profiles, distribution companies are able to target specific customer types in an attempt to change their power usage behavior. The decomposition is what this paper aims at achieving where the results produced can be utilized by distribution companies to achieve their desired outcomes.

A complete methodology has been developed in this paper to analyze and extract relevant information from unknown load profiles to decompose into known sector/sub-sector types. The developed technique utilizes the Euclidean distance and K-means clustering in the decomposition process. The analysis is completed on a month by month basis over a selected time frame accounting for seasonal change to provide the most accurate results. This paper presents the complete approach and procedure in determining the breakdown of a set of unknown profiles.

II. Data Collection

The data in this paper is obtained from an Australian distribution company (ENERGEX) and incorporates 59 load profiles spanning a period of approximately four years from the beginning of 2007 to the end of 2010. The data collected contains power consumption drawn by distribution feeders in 30 minute intervals totalling 48 points for each day, throughout the duration of the period of study. The scope of ENERGEX's distribution system is vast, with a total of 1,149 11kV feeders in its network. The study consisted of 59 feeders from a network of 1,149, in which the data was requested to be captured with a range of customer demographics. The

Ghavameddin Nourbakhsh (g.nourbakhsh@qut.edu.au), Gary Eden (gary.eden@connect.qut.edu.au), Dylan McVeigh (dylan.mcveigh @connect.qut.com.au) and Arindam Ghosh(a.ghosh@qut.edu.au) are with the Queensland University of Technology (QUT), Brisbane 4001, Australia.

methods developed in this paper can easily accommodate an increased scope of data in its analysis. Increased data in the analysis would also increase the accuracy of the results.

The data sets obtained consist of what is originally assumed to be 20 residential, 18 commercial, 16 industrial load profiles and an additional 5 unknown feeder load profiles. To provide accurate comparison between the known and unknown profiles, the known load profiles are required to meet multiple criteria to be included in the model:

1. The load profiles must have the same timeframe and duration.
2. During the timeframe for analysis, network topography and customer sector proportions should not dramatically change.
3. The load profiles must be composed of a majority of one major load sector.

The unknown load profiles only need to meet the first requirement, to capture the same seasonal changes present in the known load profiles.

III. DATA PREPARATION

To provide meaningful analysis between each profile, the profiles are required to be converted into a comparable form. This is done by converting the half hourly Power consumption into a Per Unit (PU) measurement, using (1).

$$PU\ Half\ Hourly = \frac{Half\ Hour\ Usage}{Total\ Daily\ Usage} \qquad (1)$$

Before this process was undertaken, data anomalies were processed into an acceptable range, where power outages or data corruption may have caused the data to skew. The abnormal data points were replaced with the half hourly power consumptions from the previous week. Throughout the paper, the per-unitized load profiles are stated as the profiles for simplicity.

As the analysis was performed on a month by month basis, an entire month of data needed to be represented in a simpler representation. This was achieved by representing an entire month of data into key characteristic periods of weekday, Saturday and Sunday. The weekdays themselves had minimal variation during the week and were averaged into a single day to place greater emphasis on the weekend sections. The weekend sections were shown to have characteristically different profiles between the sectors and were chosen to help discriminate between the sectors during the analysis. Each day of the month was then averaged into their respective categories.. Completion of this process forms the final profile representing a single month. Figure 1 is an example of a residential profile in June with indication of which each section represents.

The time period of study is selected from the beginning of 2008 to the end of 2009, a two year period. This period was selected as this fulfils criterion 1 regarding data selection, indicated before. To test the integrity of the selected data, load profiles were individually tested by comparing one month of the first year to the same month of the following year. This was to ensure criterion 2 was fulfilled and no major changes to the usage pattern of power occurred over the two year period. To test for changes, the load profiles were divided by each other and graphed to display the change each data point made.

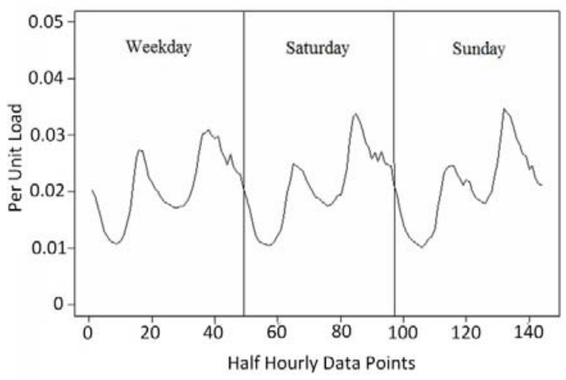

Fig. 1 A single feeder load profile, a weekday and weekends representing a month

Figure 2 is an example of the graphs produced with each line representing the change of a data point in a profile compared to that same data point in a profile of the same month in the previous year. Thresholds of 10% were assigned to ensure consistent validation occurs between profiles. Figure 3 clearly shows major changes in the power usage and this profile was not used in the two year analysis.

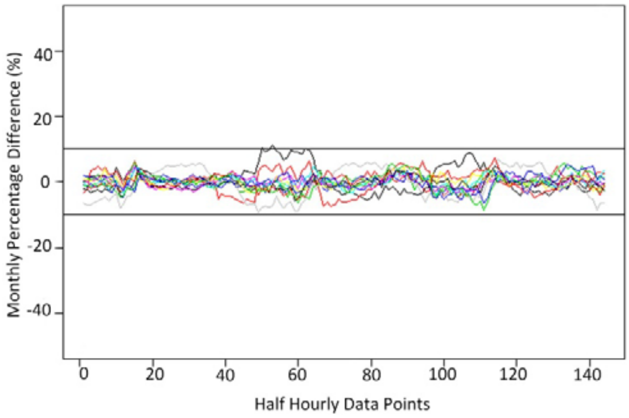

Fig. 2 Accepted, monthly load profile comparison

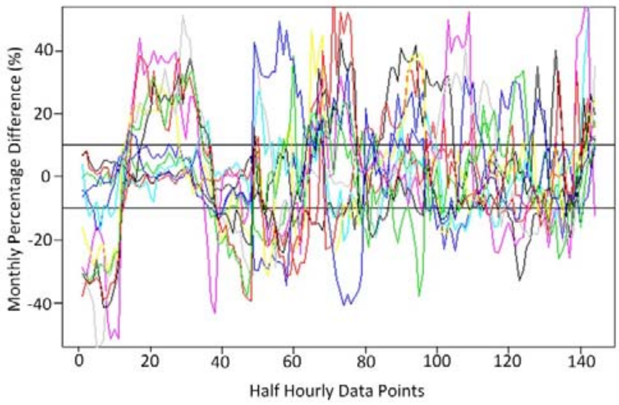

Fig. 3 Rejected, monthly load profile comparison

The basis of the Euclidian distance method utilizes the distance between profiles to calculate how well a profile is matched to another. The fundamental basis of this method is used to check the initial assumptions of which profiles are based in which sectors. The Euclidian distance is derived from the Pythagorean formula stated in (2).

$$Distance = \sqrt{(a_1 - b_1)^2 + \cdots + (a_{144} - b_{144})^2} \qquad (2)$$

where a and b are the profiles we are calculating the distance between them, with each profile having 144 data points.

Once the distances are calculated between a profile and every other profile, graphs are produced displaying the distances. For each feeder to be used in the analysis, a graph such as Figure 4 is produced and used to determine if a profile is correctly assigned. Results obtained indicate a minority of original profile classifications are incorrect. Reassignment is performed on these incorrectly classified profiles before clustering and decomposition commences. Figures 4 and 5 provide a visual aid in showing changes made regarding incorrect classification. In the example provided in Figure 4, the industrial profile can be seen to have higher correlation to other industrial profiles. Note that the smaller the value, the closer the profile is to another. Other profiles such as the bottom residential line in Figure 4 can be seen to be incorrectly classified as residential and was reassigned to industrial in Figure 5.

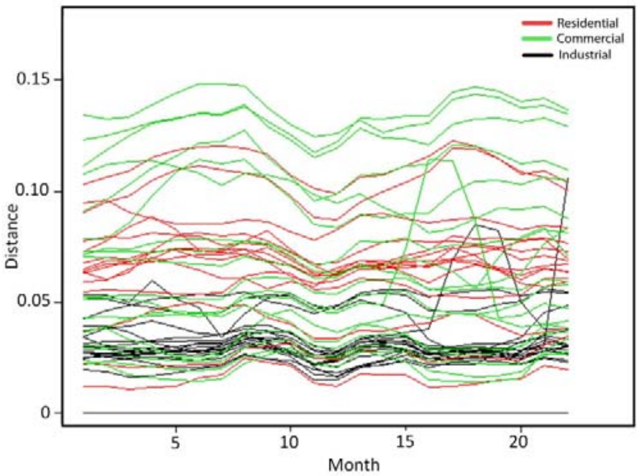

Fig. 4 Euclidean distance graph before re-assignment

Using the Euclidian distance to check the classification assumption contributes to the clustering method, as incorrectly classified profiles may form their own cluster in the wrong sector. This would alter the results significantly when decomposing the unknown load profile.

A comparison of Figure 6 and 7 shows an incorrect grouping of data in Figure 7, where it is evident that a cluster in commercial profile is actually residential. The green profiles in the commercial profile image of Figure 7 show strong correlation to all profiles in the residential profile image of Figure 6. The Euclidian distance analysis ensures the incorrectly assigned profiles are reassigned.

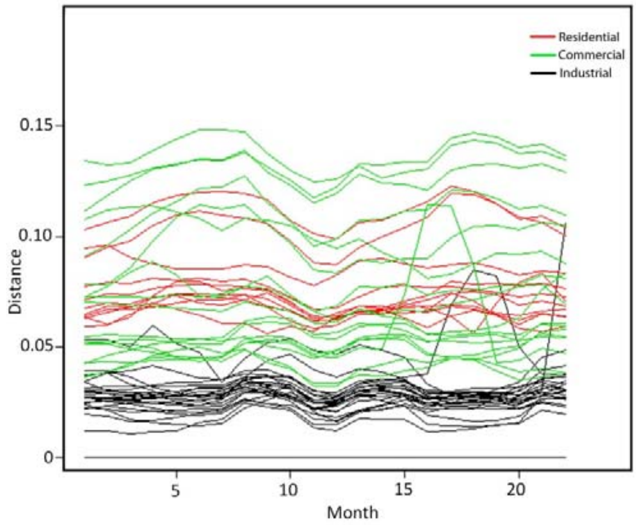

Fig. 5 Euclidean distance graph after re-assignment

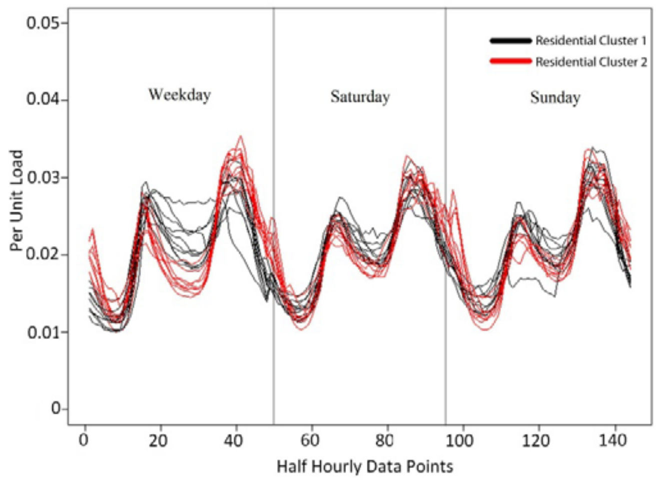

Fig. 6 Residential Profiles displaying clustering into two groups

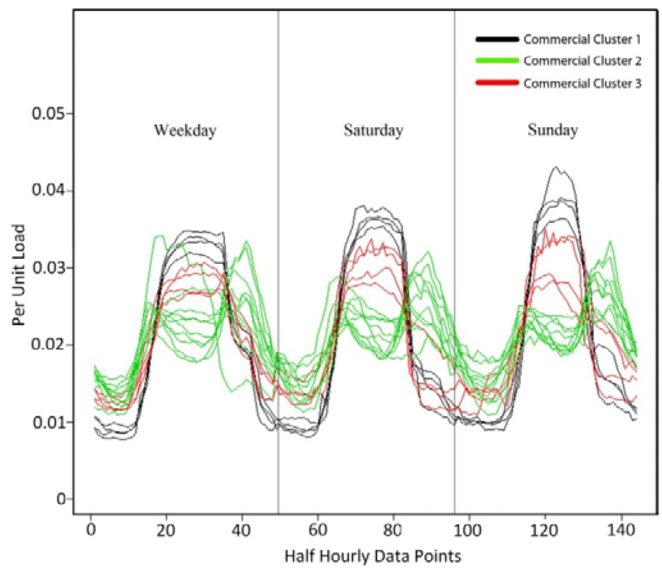

Fig. 7 Incorrectly classified Commercial profiles which form their own cluster

IV. Methodologies

Clustering and Euclidian Distance approaches are used to analyze the known and unknown profiles. Before the final approach was developed using both methodologies, both were used individually to analyze the data. They were then combined into a single method to produce optimal results.

As each profile differs slightly, different assumptions were made to account for their differences. Clustering assumes that there are sub-sectors within each main sector reflecting differences, for instance, in socio-economic classes. The slight variations of profiles with each sub-sector are then averaged to form the mean signature profile of that sector. These signature profiles are used to decompose unknown profiles.

The Euclidian distance method assumes that the majority of the differences between each profile are accounted for by profiles having contributing factors of other sectors/sub-sectors. This assumption is made as a power line or transformer may never supply to just one sector/sub-sector. By comparing the distances between each of the profiles, each is re-assigned a new proportion breakdown in terms of the three distinct sectors, or their sub-sectors.

The developed methodology utilized both the Euclidean distances and clustering to support each other. Clustering used the Euclidian distance results to ensure profiles where correctly assigned and to cross validate the initial assumptions. The Euclidian distance method uses the output of the clustering analysis to compare subsectors and their mixture in load profiles.

V. Euclidian Distance

The Euclidian distance methodology was derived by analyzing how close vector profiles are from each other. This section of the paper describes the method development and the results produced

The Euclidian distance is calculated from each known profile to every other known profile. Figure 8 is an example of a graph generated for the 2 year period. In this figure, an assumed industrial profile (red) is mostly composed of other industrial profiles; however it has some influence from commercial profiles (green). This indicates that some profiles are not purely comprised of a single sector.

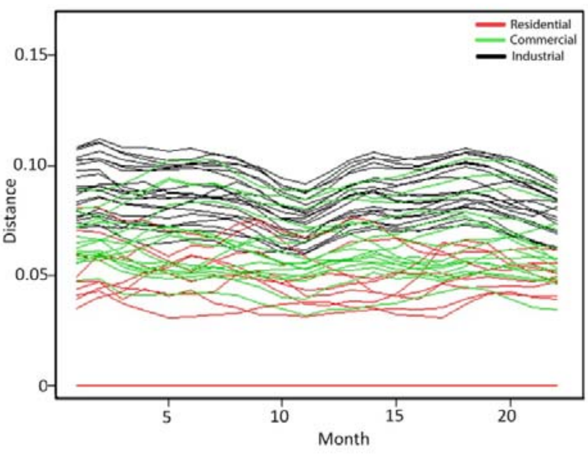

Fig. 8 Euclidean distance comparison of load profiles

To account for this, each profile is assigned a proportion breakdown of the sub-sectors before being compared to the unknown profile. The process of determining the number of sub-sectors and how each profile is assigned to the sectors is described in the next section. The Euclidian distance procedure is as follows:

A matrix of distances is created for each month of data and each distance is scaled so that closer profiles have a much larger influence. The scaling applied is simply raising the Euclidian distance in each profile by the power of four, as in (3). In the development of this procedure, this was shown to provide a higher accuracy of decomposition and greater discrimination between vastly different profiles over lower scaling powers.

$$D_{a,b} = \left(\sqrt{(a_1 - b_1)^2 + \cdots + (a_{144} - b_{144})^2}\right)^4 \quad (3)$$

A distance ($\boldsymbol{D}$) matrix is created, as in (4), where n is the number of known profiles; 59 in this study.

$$\boldsymbol{D} = \begin{bmatrix} D_{1,1} & D_{1,2} & \cdots & D_{1,n} \\ D_{2,1} & D_{2,2} & \cdots & D_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n,1} & D_{n,2} & \cdots & D_{n,n} \end{bmatrix} \quad (4)$$

To re-assign the proportions of each profile, the scaled distances are inverted and the sum of value for each profile is linearly scaled back to one, as given in (5). This will force smaller distances to have much higher values than others. The diagonals are set to zero as the decomposition of a profile should not be composed of the same profile. This is the basis of take one out cross validation [9], which has been applied as a method of self-checking for clustering and initial assumptions.

$$Scaled\ \boldsymbol{D} = \begin{bmatrix} 0 & \frac{k_1}{D_{1,2}} & \cdots & \frac{k_1}{D_{1n}} \\ \frac{k_2}{D_{2,1}} & 0 & \cdots & \frac{k_2}{D_{2,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{k_n}{D_{n,1}} & \frac{k_n}{D_{n,2}} & \cdots & 0 \end{bmatrix} \quad (5)$$

The k values are calculated to scale the sum of each row to 1, by calculating k as in (6).

$$k_i = \frac{1}{\sum_{j=1}^{n} \frac{1}{D_{i,j}}}; j \neq i \quad (6)$$

The resulting values are multiplied by a matrix consisting of the original assumed profile sectors as shown in (7). This will result in new proportions of each sector for each profile, where s is the number of sectors/sub-sectors and n is the number of profiles; in this case eight sectors for all 59 profiles. Scaling to one by using the k_i values ensures that the sum of the new proportions for each profile equates to one.

$$\begin{bmatrix} 0 & \frac{k_1}{D_{1,2}} & \cdots & \frac{k_1}{D_{1n}} \\ \frac{k_2}{D_{2,1}} & 0 & \cdots & \frac{k_2}{D_{2,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{k_n}{D_{n,1}} & \frac{k_n}{D_{n,2}} & \cdots & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} P_{1,1} & P_{1,2} & P_{1,s} \\ P_{2,1} & P_{2,2} & P_{2,s} \\ \vdots & \vdots & \vdots \\ P_{n,1} & P_{n,2} & P_{n,s} \end{bmatrix} \tag{7}$$

VI. Application and Results

Eight sub-sectors were used to decompose the profiles using the distance method; two for Residential and three for Commercial and Industrial. The number of sub-sectors was chosen from the results of the clustering method.

Figure 9 is a ternary plot showing the averaged decomposition of each profile in terms of the three sectors of, Residential, Industrial and Commercial. Each sub-sector of a sector is added together to create a breakdown in terms of the three main sectors.

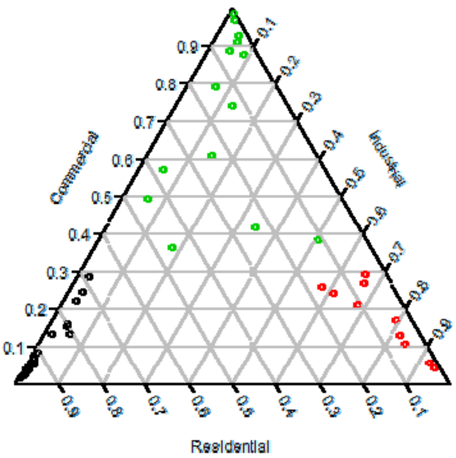

Fig. 9 Ternary plot decomposition of each known profile

As can be seen, a majority of the residential profiles are strongly composed of their own type, whilst minorities are composed of a mixture of the other sectors. A strong point to note is that this plot indicates that there is minimal residential-industrial mixture however both sectors mix with commercial. Logically this follows the existing suburban formations, as most residential units are not zoned near industrial areas whilst office buildings and other commercial business can operate in both environments.

Decomposition of the unknown profiles is done in the same manner as the proportional decomposition of the known profiles. The Euclidian distance is calculated from the unknown profile to every known profile. Inversion and quadratic scaling is applied and the unknown profile is assigned a proportion breakdown of the sectors or sub-sectors. This ensures that profiles which are similar make up the majority of the proportion.

Table I displays the averaged decomposition over the two year period in terms of the sub-sector proportions; R1 and R2 relating to residential, I1, I2 and I3 relating to industrial and C1, C2, and C3 relating to commercial.

TABLE I
UNKNOWN FEEDER DECOMPOSITION PERCENTAGE

Feeder	**R1**	**R2**	**I1**	**I2**	**I3**	**C1**	**C2**	**C3**
1	39.9	44.4	1.9	1.8	0.9	4.5	5.6	1.0
2	27.5	52.7	1.5	0.8	0.7	8.3	7.2	1.3
3	42.2	49.0	0.9	0.5	0.4	3.1	3.1	0.8
4	29.0	51.4	2.3	1.3	1.0	6.7	7.2	1.1
5	9.0	17.2	8.1	3.9	3.4	40.2	15.9	2.3

Figure 10 displays the time variance of the decomposition of unknown profile 2. As can be seen the decomposition varies from a high residential breakdown to a high residential, medium commercial breakdown. This displays how much variation can be accounted for by the respective customer sectors due to season influences.

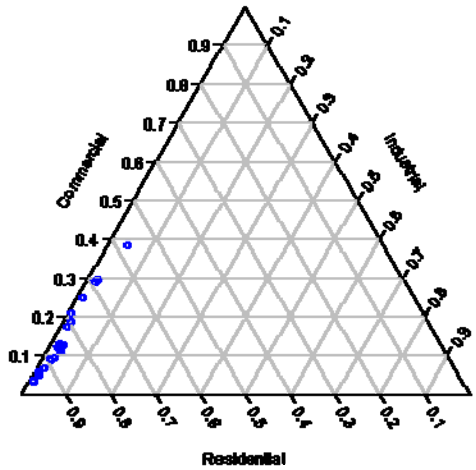

Fig. 10 Displaying the variance of decompositions over time of an unknown profile

To compare the methods and their accuracy, profiles are recreated and compared to the actual profile. A predicted profile can be created by multiplying the contributions of the known sub-sectors with their respective profiles and then summing the results, essentially recreating the unknown profile for each particular month. The SSE is then calculated for comparison.

VII. Clustering

While the previous method looks at the distance of profiles and uses a large data set for decomposition, the clustering method looks at creating a set of signature profiles for each sub-sector. These signature profiles are then passed through a Minimize SSE function for decomposition. To create these signature profiles, K-means clustering is applied

K-means clustering [10] is used to group similar sets of profiles together. Like other clustering methods, K-means uses the Euclidean distance to determine if vector profiles are similar to others. For this clustering method to produce the most accurate results, the optimal number of clusters for each sector needs to be determined.

Calculation of the optimal number of clusters is performed using silhouettes, which shows how close an assigned profile in a certain cluster is to other clusters. Using the averaged monthly values shows how appropriately the data has been clustered. An optimal number of clusters can then be determined by comparison of this averaged value across a range of different cluster values.

Silhouette calculations are done via (8), where for each profile, i, $a(i)$ is the average distance to other profiles in the same cluster and $b(i)$ the average distance to profiles of the nearest cluster. It returns a value between -1 and 1 representing the ratio of how close a profile is to its assigned cluster compared to the next closest cluster. Values closer to 1 indicate the profile is very closely positioned to its assigned cluster. A value of zero indicates that the profile is at equal distance between clusters and a value of -1 indicates the profile is closer to a different cluster than the one it is assigned to.

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (8)$$

This analysis was completed on a month by month basis with the monthly results of each particular cluster averaged over the 24 month period. The values are tabulated in Table II.

TABLE II
INTERPRETATION OF SILHOUETTE ANALYSIS

	Number of clusters			
	2	**3**	**4**	**5**
Residential	0.782508	0.769428	0.596708	0.713344
Industrial	0.598946	0.792220	0.720742	0.716830
Commercial	0.838026	0.842235	0.799648	0.771778

The results indicate that the optimal number of clusters for each sector is 2, 3 and 3 for Residential, Commercial and Industrial, respectively. These values correspond to the largest averaged silhouette value. Each profile is clustered into their respective sub-sector for each month of the two year period of interest.

The profiles within each cluster are then averaged into a single signature profile. This forms the basis of each sub-cluster, which is then used to decompose the unknown feeder. The procedure used to decompose the unknown profile aims at reducing SSE in the decomposition. This is completed via minimizing (9).

$$\frac{1}{n}\sum_{i=1}^{n}\left|\left|y_i - \left(\sum_{j=1}^{m} X_{i,j}\, p_{j,1}\right)\right|\right|^2 + \lambda\, \mathrm{J}_2(f) \quad (9)$$

Where; y_i is the actual unknown profile, X is the matrix of signature profiles for a specific month, $\mathrm{J}_2(f)$ is the penalty on the roughness, and λ is the smoothing parameter. This equation is minimized to find the best fit for p, the vector of contributions, giving the random feeder contributions for a specific month. This minimization is then repeated for each month in the study, allowing for month-by-month contributions to be calculated.

Figure 11 shows a predicted (red) vs. actual (black) plot of an unknown profile using the penalized least squares regression technique [11].

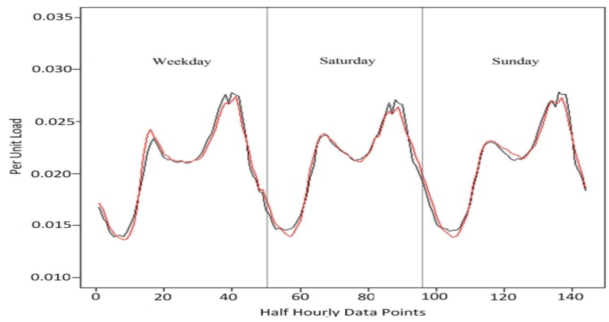

Fig. 11 Predicted vs. Actual profile for residential in October 2009

This procedure minimizes the SSE by choosing the optimal linear combination of signature profiles. This may overlook some highly correlated profiles in favor of others which have a better combination to reduce the SSE.

Table III gives decomposed proportions of each unknown feeder, breaking up into eight sub-sectors, over the two years.

TABLE III
UNKNOWN FEEDER DECOMPOSITION PERCENTAGE

Feeder	**R1**	**R2**	**I1**	**I2**	**I3**	**C1**	**C2**	**C3**
1	47.5	39.2	1.8	3.4	0.9	1.2	1.4	4.6
2	20.9	64.9	3.9	6.2	3.4	0.0	0.0	0.7
3	52.6	42.8	0.2	0.4	1.2	0.0	0.5	2.3
4	31.3	55.4	0.0	0.0	0.8	4.9	3.9	3.7
5	2.4	5.6	5.3	16.7	64.3	4.3	0.6	0.8

VIII. METHOD COMPARISON

The aim of the model developments was to accurately decompose unknown profiles into known sectors or sub-sectors. Both methods have shown compatible results and each has strengths and weaknesses. The compared results of the decompositions are displayed in Table IV. The results are shown to be similar in most cases where unknown profiles 2 and 5 have a discrepancy in the Industrial and Commercial components.

The Euclidian distance analysis showed that unknown profiles 2 and 5 were closer to commercial in their shape.

TABLE IV
DECOMPOSITION COMPARISON

Feeder	**Method**	**Res**	**Ind**	**Com**
1	Clustering	86.7%	6.1%	7.2%
	Distances	84.3%	4.6%	11.1%
2	Clustering	85.8%	13.5%	0.7%
	Distances	80.2%	3.0%	16.8%
3	Clustering	95.4%	1.8%	2.8%
	Distances	91.2%	1.8%	7.0%
4	Clustering	86.7%	0.8%	12.5%
	Distances	80.4%	4.6%	15.0%
5	Clustering	8.0%	86.3%	5.7%
	Distances	26.2%	15.4%	58.4%

Further comparison between both approaches was made by comparing the SSE of the predicted vs. actual values. Table V displays the comparison of SSEs.

TABLE V
SSE COMPARISON

Feeder	Distance	Clusters
1	0.0159	0.0105
2	0.0071	0.0015
3	0.0063	0.0026
4	0.0093	0.0051
5	0.0086	0.0034

The results show that clustering technique has the smallest SSE in all cases. This can be directly contributed to the decomposition method used, which aims to minimise the error.

Figure 12 shows the comparison of an actual profile (black) to a predicted profile using an input of 8 sub-sectors to decompose the profile (red) and the clustering method (green). Inspection of Figure 12 shows the green line (clustering) is consistently closer to black line (actual) over the entire graph, which is expected due to the lower SSE.

Not only does the clustering method produce the smallest SSE in decomposition, the data requirements for using this method are also smaller. Once signature profiles have been created, they can be applied to other systems for decomposition. This has the benefit for power systems which are smaller in nature.

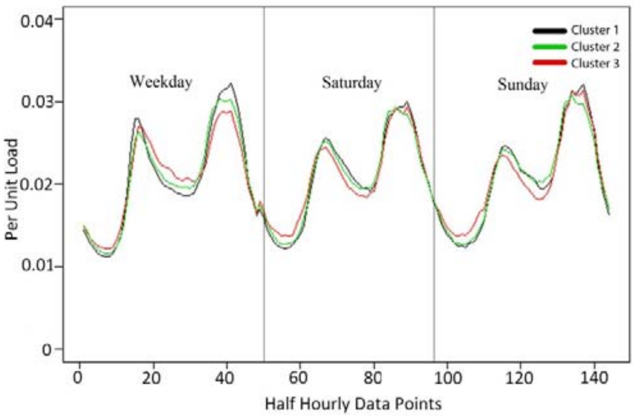

Fig. 12 Comparison of methodologies

In underdeveloped countries, the planning criteria for residential areas can differ compared with that of developed countries, so residential zones may occur near heavy industry and could be supplied through the same power lines. This would create highly mixed sectors and new profiles which the Euclidian distance method may not encapsulate. The Euclidian distance method is based on prior knowledge of various mixture types and the addition of a profile which differs greatly may produce incorrect decomposition. The clustering method accounts for this issue as the decomposition will choose signature profiles to create the unknown profile in an accurate way.

If the clustering method is used, the initial sector classifications need to be fairly accurate in order to achieve the correct decomposition. If the initial classifications are not correct, results obtained will not reflect a true decomposition of an unknown feeder. Hence, the Euclidean distance is useful in determining correlation patterns between sets of load profiles, to cross validate the initial assumptions of sector classification. If large data sets can be provided which cover a wide range of sector mixtures, the Euclidian distance method could prove more accurate when decomposing unknown profiles. Overall, the limitations of this method are based on having sufficient data, and that the unknown profiles do not differ greatly from the data set.

IX. CONCLUSION

This paper proposed techniques and procedures to decompose unknown chronological load profiles into known sector/sub-sector profiles. K-means clustering and Euclidean distance techniques were used as part of the procedures developed in this paper to decompose unknown load profiles. The procedures were designed for chronological loads to capture the changes in customer load consumption due to new developments and seasonal changes.

The method developed in this paper was shown to be accurate in decomposing unknown load profiles, in cases provided by an Australian distribution company. Suitable results were obtained when both techniques were utilised together, by using the Euclidian distance technique to cross validate and reassign the initial assumptions of profile classifications before being applied to the clustering analysis.

ACKNOWLEDGEMENT

The authors wish to thank ENERGEX Limited of Australia for providing the feeders data.

REFERENCES

[1] A. Mutanen, M. Ruska, S. Repo, P. Jarventausta, “Customer classification and load profiling method for distribution systems,” IEEE Transactions on Power Delivery, vol. 26, pp. 1755-1763, Jul. 2011.

[2] N. Anuar, Z. Zakaria, “Cluster validity analysis for electricity load profiling,” in Proc. 2010 IEEE International Conference on Power and Energy, Kuala Lumpur, Malaysia, 2010, pp. 35-38

[3] Z. Zakaria, K. L. Lo, “Two-stage fuzzy clustering approach for load profiling,” in Proc. 44th International Universities Power Engineering Conference (UPEC), Shah Alam, Malaysia, 2009, pp. 1-5.

[4] Z. Zakaria, M. N. Othman, M. H. Sohod, “Consumer load profiling using fuzzy clustering and statistical approach,” in Proc.4th Student Conf. Res. Develop., Selangor, Malaysia, 2006, pp. 270-274.

[5] B. D. Pitt, D. S. Kitschen, “Application of data mining techniques to load profiling,” in Proc. 21st IEEE International Conference on Power Industry Computer Applications, Manchester, UK, 1999, pp. 131-139.

[6] A. H. Nizar, Z. Y. Dong, and Y. Yang "Power utility nontechnical loss analysis with extreme learning machine method,” IEEE Transactions on Power Systems, vol. 23, pp. 946-955, Aug. 2008.

[7] N. Yamaguchi, J. Han, G. Ghatikar, “Regression models for demand reduction based on cluster analysis of load profiles,” in Proc. Sustainable Alternative Energy (SAE)., 2010, pp. 1-7.

[8] Kang, M. S., C. S. Chen, Y. L. Ke, and T. E. Lee, “Stochastic load flow analysis by considering temperature sensitivity of customer power consumption,” in Proc. IEEE Bologna Power Tech Conference, Bologna, Italy, 2003.

[9] M. Stone, “Cross-validatory choice and assessment of statistical predictions”, Journal of the Royal Statistical Society, Series B (Methodological) Vol. 36, pp. 111-147, 1974.

[10] D. T. Pham, S. S. Dimov, C. D. Nguyen, “Selection of K in K-means clustering,” Proc. Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, pp. 103-119, 2005.

[11] S. Wood. (2002) Package ‘mgcv’. [Online]. Available: http://cran.r-project.org/web/packages/mgcv/mgcv.pdf

Ghavameddin Nourbakhsh received the B.Sc. and M.Sc. in electrical engineering from California State University in Sacramento, in 1980 and 1981, respectively, the M.Sc. research degree in power system reliability from the University of Saskatchewan, Saskatoon, SK, Canada, in 1991, and the Ph.D. degree in electrical engineering and computer science from Queensland University of Technology, Brisbane, Australia, in 2011.

He worked for power supply authorities in Iran before obtaining the M.Sc. research degree. In 1994, he joined the academic staff of the School of Electrical Engineering and Computer Science, Queensland University of Technology. His interests are in Power System Reliability and Cost/Worth analysis and Smart Grid Operations.

Gary Eden was born on October 14, 1990 in Australia. He is currently pursuing the B.Eng. undergraduate degree in electrical engineering and mathematics at the Queensland University of Technology, Brisbane, Australia.

His interests lie in power system operation and control.

Dylan McVeigh was born on November 23, 1990 in Australia. He is currently pursuing the B.Eng. double undergraduate degrees in electrical engineering and mathematics.

Mr. McVeigh is a recipient of the Australian Power Institute (API) Bursary and works at the Australian Energy Market Operator (AEMO), developing programs to improve accuracy and efficiency of electricity transmission and market studies.

Arindam Ghosh (S’80, M’83, SM’93, F’06) received the Ph.D. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 1983.

Currently, he is the Professor of Power Engineering, Queensland University of Technology (QUT), Brisbane, Australia. Prior to joining the QUT in 2006, he was with the Dept. of Electrical Engineering, Indian Institute of Technology Kanpur, for 21 years. His interests are control of power systems and power-electronic devices. Prof. Ghosh is a Fellow of Indian National Academy of Engineering (INAE) and IEEE.